AP Statistics Name: ______________________ Date: ________

Regression Analysis

Does how long young children remain at the lunch table help predict how much they eat? Here are data on 20 toddlers observed over several months at a nursery school. "Time" is the average number of minutes a child spent at the table when lunch was served. "Calories" is the average number of calories the child consumed during lunch, calculated from careful observation of what the child ate each day.

Time	21.4	30.8	37.7	33.5	32.8	39.5	22.8	34.1	33.9	43.8
Calories	472	498	465	456	423	437	508	431	479	454
Time	42.4	43.1	29.2	31.3	28.6	32.9	30.6	35.1	33.0	43.7
Calories	450	410	504	437	489	436	480	439	444	408

1. Describe the relationship in a graph and by regression analysis.

There is a moderately strong negative linear relationship between calories eaten and time spent at the table for these toddlers.

Regression Analysis: Calories versus Time

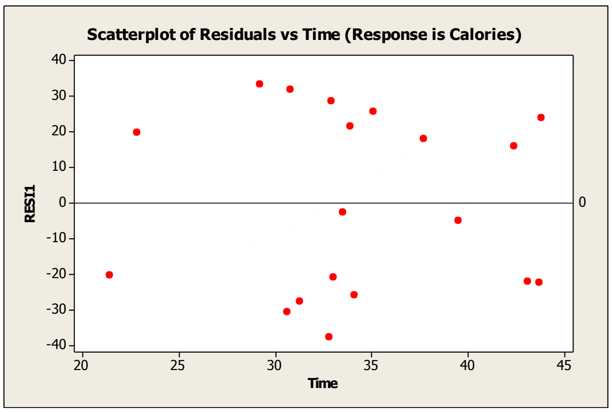
```
The regression equation is
Calories = 561 - 3.08 Time


Predictor      Coef  SE Coef      T      P
Constant     560.65    29.37  19.09  0.000
Time        -3.0771   0.8498  -3.62  0.002


S = 23.3980   R-Sq = 42.1%   R-Sq(adj) = 38.9%
```

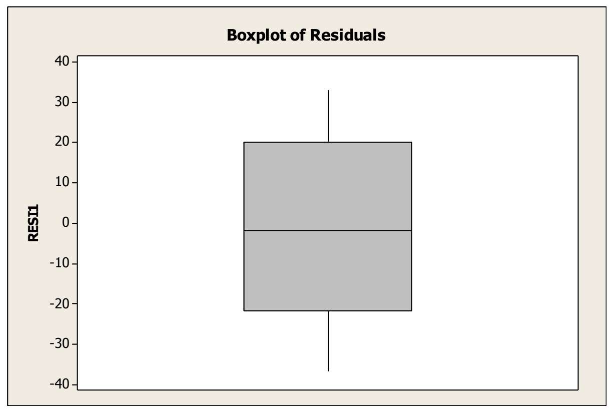
From the regression analysis we see that the correlation $r = -\sqrt{0.421} = -0.649$ which is consistent with the graph. (We take the negative root because the slope of the regression line is negative.)

2. Perform a test of significance for these data. Use the back of this sheet if you need more space.

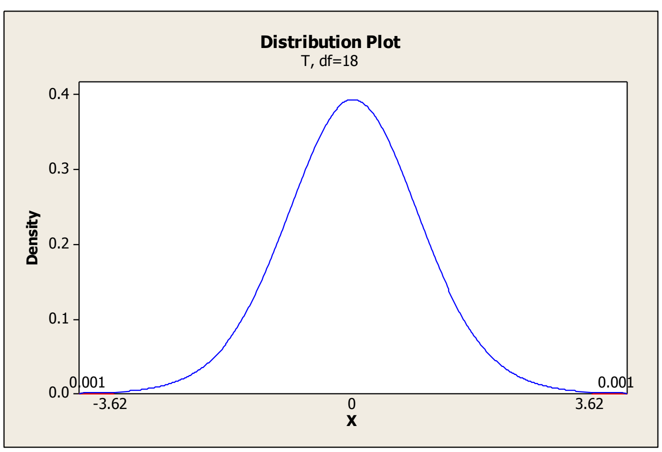
State: $H_0: \beta = 0$ Where β is the true slope of the regression line of calories on time at table
$H_a: \beta \neq 0$ for these toddlers.
Choose significance level $\alpha = 0.05$

Plan: If conditions for inference are met, conduct a linear regression *t* test.
We must assume these months of observation are a random sample of the toddlers' eating habits in the nursery school and that the amount they eat at the table is independent of each other. Looking at a scatterplot of residuals vs. time, we can see no obvious violations of the conditions for linear fit and equal variance. (There is a patternless scatter, and the spread of calories eaten is roughly the same for all lengths of time.) A boxplot of the residuals shows no outliers or obvious skew, so we can assume the Normality condition is satisfied.





Do: The output gives us the test statistic $t = \frac{b-0}{SE_b} = -3.62$
This test statistic follows a *t* distribution with $n - 2 = 20 - 2 = 18$ degrees of freedom.



The P-value is $P(|T| > 3.62) = 0.002$ which is much less than our significance level of 0.05.

Conclude:
We reject H_0. We have strong evidence of a relationship between time spent at the table and calories consumed for toddlers at this nursery.

3. Construct a 90% confidence interval to estimate how rapidly calories consumed changes as time at the table increases.

We are looking to estimate β, the true slope of the regression line of calories on time for these toddlers, with 90% confidence. Conditions for inference were checked above and met.

The confidence interval is given by

$$b \pm t^* \cdot SE_b$$

The critical value t^* for a 90% confidence interval with $df = 18$ is 1.734. (Use t distribution table or "invt" function on calculator).

$$-3.0771 \pm 1.734(0.8498)$$
$$-3.0771 \pm 1.4736$$
$$(-4.551, -1.603)$$

We are 90% confident that the interval from -4.551 to -1.603 captures the true slope of the regression line of calories consumed on time spent at the table.